

# ANDREW CHU

647-966-2880 | [contact@andrewchu.ca](mailto:contact@andrewchu.ca) | [/in/andrewchu-ca](https://in.linkedin.com/in/andrewchu-ca) | [github.com/andrewchu16](https://github.com/andrewchu16) | [andrewchu.ca](https://andrewchu.ca)

## EDUCATION

### University of Waterloo

Bachelor of Computer Science - Dean's Honours List

Apr. 2028

Waterloo, ON

## EXPERIENCE

### Software Engineer

May 2025 - Aug. 2025

CGI

Markham, ON

- Developed enterprise RAG chatbot to speed up information retrieval from **500+** internal production guides, reducing team incident response times by **17%**
- Engineered unified dashboard for remotely executing jobs on **400+** applications and servers, automating health checks, server restarts, and email notification services
- Increased chat response accuracy by **38%** through PEFT fine-tuning a Llama3 model with QLoRA

### Technical Lead (Part Time)

May. 2025 - Aug. 2025

University of Waterloo's Data Science Club

Waterloo, ON

- Spearheaded user check-in and application portal projects written in Next.js and Express with **1000+** users
- Resolved **15+** development tickets spanning frontend UI enhancements, application submission workflows, and backend database migration
- Implemented email confirmation for password resets and user portal, improving user experience and security

### Founding Engineer

May. 2025 - Present

Villio AI

Waterloo, ON

- Architected FastAPI backend for multi-tenant city services app, integrating Firebase Auth for user authentication
- Engineered agentic pipeline using LangChain for filling out forms through a chat interface for ease of use
- Deployed backend securely through Azure App Service, enabling high availability across multiple cities

### Data Engineer Intern

Aug. 2024 - Sep. 2024

Alljoined

Toronto, ON

- Debugged critical timing issue in EEG data preprocessing by implementing data logging system in Python
- Improved data collection script performance by **20x** by refactoring WebSocket script to prevent deadlocking in network communications between EEG headset and compute cluster
- Designed data processing pipeline for denoising and artifact correction using NumPy, Pandas, and scikit-learn

### Full-Stack Developer

Sep. 2022 - Jan. 2023

Digitera Interactive

Ottawa, ON

- Developed "Skule News" project, a mobile school news app to streamline student communication written in Flutter
- Integrated Firebase for authentication, real-time updates, and cloud file storage for secure backend support
- Reduced frequency of in-app crashes by **20%** by resolving **6+** bugs resulting from incorrect Firebase queries

## PROJECTS

### [chat.andrewchu.ca](https://chat.andrewchu.ca) - Personal AI Assistant | Transformers, Nginx, PyTorch, Docker, Postgres, Next, FastAPI

- Built a personal AI chatbot fine-tuned on my personal experiences using QLoRA, hosted on a Raspberry Pi
- Cut LLM inference times by **67%** by embedding queries with EmbeddingGemma and caching LLM answers with vector similarity search

### Tinyraytracer - Computer Graphics Renderer | C++, Go

- Developed a fast and simple raytracer written in **Go** and **C++** for **multithreaded** CPU-based 3D rendering
- Accelerated I/O image write operations **8x** by leveraging Goroutines to write **concurrent**, performant code
- Implemented effects including specular lighting, reflection, and refraction using the Phong reflection model

## TECHNICAL SKILLS

**Languages:** TypeScript, Python, C/C++, Java, Go, C#, CSS, HTML, Bash

**Databases:** PostgreSQL MongoDB, MySQL, Supabase, Cosmos DB

**Frameworks:** Next.js, React, Express, Flask, FastAPI, Jest, SvelteKit, Electron, LangChain, nginx

**Developer Tools:** Azure, WebLogic, Linux, Git, Figma, Docker, Postman, Google Cloud Platform